Journal of
Modern Medical Information Sciences

Summer 2022. Vol 8. Issue 2

Research Paper

Predicting Gestational Diabetes Using an Intelligent Algorithm Based on Artificial Neural Network

***Mohammadjavad Hosseinpoor**[1]

1. Department of Computer Engineering, Estahban Branch, Islamic Azad University, Estahban, Iran.

Use your device to scan and read the article online

Citation Hosseinpoor M. [Predicting Gestational Diabetes Using an Intelligent Algorithm Based on Artificial Neural Network (Persian)]. Journal of Modern Medical Information Sciences. 2022; 8(2):126-139. https://doi.org/10.32598/JMIS.8.2.3

https://doi.org/10.32598/JMIS.8.2.3

CC BY NC

Article Info:

Received: 16 Aug 2021

Accepted: 11 Apr 2022

Available Online: 01 Jul 2022

Key words:
Gestational diabetes, Genetic algorithm, Artificial neural network

ABSTRACT

Objective Due to the large amount of data for people with diabetes, it is very difficult to extract the predictors of diabetes. Data mining science can discover the predictors of diseases and help physicians and medical staff in predicting and diagnosing diseases.

Methods This is an applied survey study conducted in 2020 using the dataset used by Mirsharif et al. The study population includes 105 cases with data registered from 2011 to 2014 in a specialized women's medical center in Tehran, of which 80 were for healthy women and 25 were for women with gestational diabetes. MATLAB software was used to analyze and evaluate the results.

Results The results and comparisons showed the high efficiency of the proposed method in predicting gestational diabetes. The accuracy of the proposed method was 93%, which was more accurate than the method proposed by Mirsharif et al.

Conclusion The proposed prediction method has good performance and high accuracy compared to previous methods. Therefore, this intelligent and unsupervised method can be used to predict gestational diabetes.

Corresponding Author:
Mohammadjavad hosseinpoor, PhD.
Address: Department of Computer Engineering, Estahban Branch, Islamic Azad University, Estahban, Iran.
Tel: +98 (917) 3078032
E-mail: hosseinpoor.mohammadjavad@gmail.com

Journal of
Modern Medical Information Sciences

Extended Abstract

Introduction

Diagnosing diseases is a very challenging and complicated process for experts in medical sciences; therefore, it is necessary to use appropriate data mining methods to make a correct diagnosis in medical issues. The use of algorithms and data mining techniques can provide patterns and data regarding the factors of various diseases; based on these data, doctors and medical professionals can act for the prevention of diseases. Due to the existence of a huge amount of data about patients with diabetes, it is not be possible to extract predictors of diabetes from this huge amount of data. Data mining science can discover the predictors of diseases and help doctors and medical staff in the prediction and diagnosis of diseases.

Methods

The is an applied survey study conducted in 2020 using the dataset of Mirsharif et al.'s study. The study population includes 105 female patient data registered from 2011 to 2014 in a specialized women's medical clinic in Tehran, of which 80 were for healthy women and 25 were for women with gestational diabetes. MATLAB software was used to analyze and check the results. The used algorithm was a hybrid artificial neural network- genetic algorithm. First, unsupervised clustering operation was performed to create two clusters of diabetic and non-diabetic women using genetic algorithm on the input dataset. After performing 100 times of execution, the best extract containing the optimal clusters was obtained. After this step, the output of the genetic algorithm was considered as the input of the artificial neural network to perform the training operation. The neural network was a feedforward neural network with 20 hidden layers. In this study, by default, 75% of the datasets were used as learning samples for system learning. Moreover, to check the accuracy of the system, 25% of the datasets were used for test and validation.

Results

The results and comparisons showed the high efficiency of the proposed method in predicting gestational diabetes. In the test dataset, 25 samples were used to evaluate the method, among which 15 were for non-diabetic women and 10 were for diabetic women. In the test of the proposed method, according to the produced confusion matrix, all 12 predictions of non-diabetic women were correctly classified. Out of 13 predictions of diabetic women, 69.2% were correctly classified and 30.8% were incorrectly classified. Out of 16 samples of non-diabetic women, 75% were correctly predicted as healthy women, while 25% were diabetic. All 9 samples of diabetic women were correctly classified as diabetic. Overall, 84% of predictions were correct and 16% were incorrect (Figure 1).

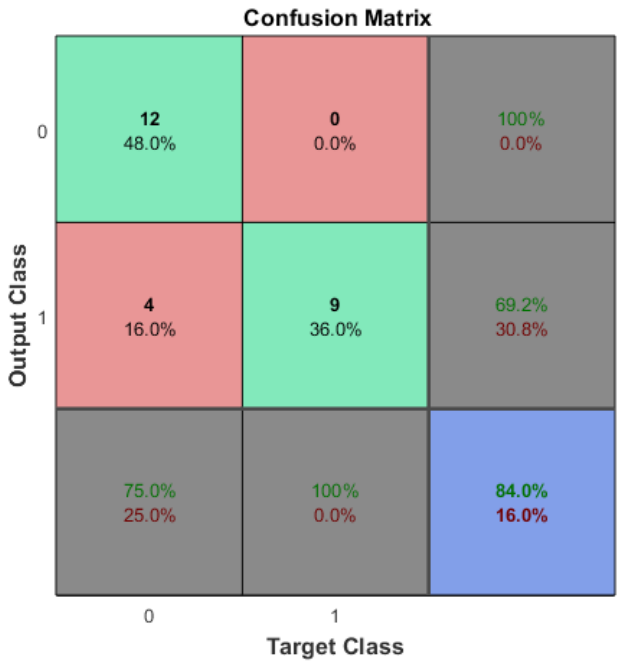


Figure 1. confusion matrix of the proposed prediction system using test data

Table 1. The results of two criteria in comparing the performance of the proposed method with Mirsharif et al.'s method

Methods	MSE	ACC
Mirsharif et al.'s method	0.0738	0.906
The proposed method	0.00823	0.932

The performance of the proposed method was compared with the method proposed by Mirsharif et al. on the same dataset based on two criteria. As can be seen in Table 1, the efficiency and performance of the proposed method was better than the Mirsharif et al.'s method in both MSE and ACC criteria. This indicates that the proposed method in unsupervised mode had the desired efficiency in facing different datasets to accurately diagnose gestational diabetes.

Discussion

In this study, an intelligent and unsupervised approach based on neural-artificial network techniques and genetic algorithm was used to predict gestational diabetes. The proposed method seeks to improve the process of differentiating women with gestational diabetes from healthy women. The neural network in the proposed method discovers hidden patterns in diabetic women compared to healthy women by examining the desired dataset. Since the function of the neural network depends on how its structure is defined, a genetic algorithm was used along with the proposed neural network whose task was to learn the structure of the neural network without supervision. The results of the implementation of the proposed system showed that the performance and accuracy of the algorithm in the desired dataset, compared to the previous similar system, were acceptable and improved significantly by 93.2%. Therefore, this intelligent and unsupervised method can be used to predict gestational diabetes.

Ethical Considerations

Compliance with ethical guidelines

This study has ethical approval from Estahban Branch, Islamic Azad University (Code: 1144819916004881397187659556).

Funding

This research did not receive any grant from funding agencies in the public, commercial, or non-profit sectors.

Conflicts of interest

The authors declared no conflict of interest.

Acknowledgements

The author would like to thank Estahban Branch, Islamic Azad University for their cooperation.

مقاله پژوهشی

پیش‌بینی دیابت بارداری با استفاده از یک الگوریتم هوشمند مبتنی بر شبکه عصبی

محمدجواد حسین‌پور[1]

۱. دانشکده مهندسی کامپیوتر، واحد استهبان، دانشگاه آزاد اسلامی، استهبان، ایران.

Use your device to scan and read the article online

Citation Hosseinpoor M. [Predicting Gestational Diabetes Using an Intelligent Algorithm Based on Artificial Neural Network (Persian)]. Journal of Modern Medical Information Sciences. 2022; 8(2):126-139. https://doi.org/10.32598/JMIS.8.2.3

doi https://doi.org/10.32598/JMIS.8.2.3

اطلاعات مقاله:

تاریخ دریافت: ۲۵ مرداد ۱۴۰۰

تاریخ پذیرش: ۲۲ فروردین ۱۴۰۱

تاریخ انتشار: ۱۰ تیر ۱۴۰۱

چکیده

هدف به دلیل وجود حجم عظیمی از داده‌ها در مورد افراد مبتلا به بیماری دیابت، امکان استخراج عوامل پیش‌بینی بیماری دیابت توسط متخصصین با استفاده از استخراج دانش از این حجم عظیم داده، امکان‌پذیر نخواهد بود. علم داده‌کاوی به کمک روش‌های مؤثر خود با هدف کشف پیش‌بینی بیماری‌ها به این مهم دست یافته و سبب کمک به پزشکان و کادر درمان در پیش‌بینی و تشخیص بیماری‌ها شده است.

روش‌ها پژوهش حاضر از نوع کاربردی‌پیمایشی در سال ۱۳۹۹ انجام شده است. در این پژوهش، از مجموعه داده میرشریف و همکاران استفاده شده است. در اینجا از روش مجموعه داده‌های اولیه برای جمع‌آوری داده‌ها استفاده شده و جامعه آماری موردنظر شامل ۱۰۵ مورد اطلاعات ثبت‌شده بیماران از سال ۱۳۹۰ تا ۱۳۹۳ در تحقیقی میدانی از کلینیک پزشکی تخصصی زنان در تهران است که از این میان، ۸۰ نفر انسان سالم و ۲۵ نفر انسان مبتلا به بیماری دیابت بارداری بودند. از نرم‌افزار متلب جهت تجزیه‌وتحلیل و بررسی نتایج استفاده شده است.

یافته‌ها نتایج و مقایسه‌های انجام‌گرفته در این پژوهش، نشان از کارایی بالای روش پیشنهادی در پیش‌بینی بیماران دیابت بارداری دارد. همچنین دقت روش پیشنهادی برابر ۹۳ درصد حاصل شد که در مقایسه با روش میرشریف و همکاران بر روی همین مجموعه داده از دقت بیشتری برخوردار بود.

نتیجه‌گیری به دلیل اینکه سیستم پیشنهادی عملکرد مطلوبی داشته و ازلحاظ دقت در مجموعه داده موردنظر نسبت به روش‌های قبلی به عدد ۹۳/۲ درصد رسیده است. پس می‌توان از رویکرد هوشمند و بدون نظارت، جهت تشخیص بیماری دیابت بارداری استفاده کرد.

کلیدواژه‌ها:

پیش‌بینی دیابت بارداری، الگوریتم هوشمند، شبکه عصبی‌مصنوعی

* **نویسنده مسئول:**

دکتر محمدجواد حسین‌پور

نشانی: استهبان، دانشگاه آزاد اسلامی، واحد استهبان، دانشکده مهندسی کامپیوتر.

تلفن: ۳۰۷۸۰۳۲ (۹۱۷) ۹۸+

پست الکترونیکی: hosseinpoor.mohammadjavad@gmail.com

مقدمه

تشخیص بیماری‌ها یک فرایند بسیار سخت و پیچیده برای افراد خبره در علوم پزشکی محسوب می‌شود، بنابراین لزوم استفاده از روش‌های مناسب داده‌کاوی جهت اتخاذ تشخیص صحیح در مسائل پزشکی ضروری به نظرمی‌رسد [۱]. درنتیجه استفاده از الگوریتم‌ها و تکنیک‌های داده‌کاوی می‌تواند الگوها و نتایجی را برای محققین فراهم کند که عوامل ابتلا به بیماری‌های مختلف را نشان دهد و براساس آن نتایج، پزشکان و دست‌اندرکاران علوم پزشکی در پیشگیری آن بیماری اقدام کنند [۲]. استخراج دانش از میان حجم انبوه داده‌ها با استفاده از فرایند داده‌کاوی می‌تواند به شناسایی قوانین حاکم بر بیماری‌ها منجر شده و اطلاعات ارزشمندی را به‌منظور شناسایی علل بیماری‌ها، تشخیص، پیش‌بینی و درمان بیماری‌ها با توجه به عوامل محیطی در اختیار متخصصان حوزه سلامت قرار دهد و این مسئله به معنای طول عمر و ایجاد آرامش برای افراد جامعه است [۳]. بنابراین ضرورت وجود روشی مناسب جهت تشخیص بیماری حس می‌شود. در این مقاله به این مهم پرداخته شده و بیماری مورد هدف آن بیماری دیابت است [۴].

یکی از مشکلات شایع دوران بارداری که در بین زنان باردار مشاهده می‌شود دیابت بارداری است [۵]. پزشکان با تکیه بر تجربیات و دانسته‌های خود، آزمایشات پیچیده و وقت‌گیر به بیماری دیابت بارداری پی می‌بردند [۶] با وجود این خطاهای انسانی اجتناب‌ناپذیر است [۷]، در این پژوهش سعی شده است با استفاده از یک الگوریتم هوشمند مبتنی بر شبکه‌های عصبی، روشی را ایجاد کنیم که به بهترین حالت شناسایی و تشخیص زودهنگام دیابت بارداری را انجام دهد. استفاده از الگوریتم‌های هوشمند می‌تواند الگوها و نتایجی را برای محققین فراهم آورد که عوامل ابتلا به بیماری دیابت بارداری را به‌سرعت تشخیص دهند [۸، ۹]. استفاده از الگوریتم هوشمند می‌تواند بهترین روش‌های درمانی و کم‌هزینه‌ترین روش ها را شناسایی کند [۱۰، ۱۱] و این امر موجب کاهش در هزینه‌های درمانی می‌شود [۱۲] در سال‌های گذشته پژوهش‌های زیادی در این زمینه صورت گرفته است. در ادامه به چندین مورد از آن‌ها اشاره می‌شود.

محمدی و همکاران، بیماری‌های مرتبط با چشم را در دیابت بررسی کردند. این بیماری آثار بدی روی چشم‌ها، سیستم عصبی و سایر اعضای بدن دارد. رتینوپاتی دیابتی عارضه‌ای ناشی از دیابت است که به دلیل تغییرات ایجادشده در عروق خونی شبکیه رخ می‌دهد. اگزودیت‌ها علامت اصلی از ریتینوپاتی دیابتی هستند. تشخیص زودهنگام می‌تواند به‌طور چشمگیری کوری را کاهش دهد. روش اتوماتیک تشخیص اگزودیت با کنتراست پایین تصاویر دیجیتال و روش خوشه‌بندی فازی برای بیماران دیابتی با مردمک چشم غیرمستمع، پیشنهاد شده است. پیش‌پردازش افزایش کنتراست قبل از ۴ ویژگی به نام شدت، شدت خطای استاندارد، رنگ و یک تعداد از پیکسل‌های لبه، پیاده‌سازی شده است. پارامترهای تقسیم‌بندی ضخیم با استفاده از روش خوشه‌بندی فازی استخراج شده‌اند [۱۳]

زاهدی‌فرد و همکاران، با استفاده از داده‌کاوی و تمرکز بر روی داده‌های مربوط به افراد دیابتی ثبت‌شده در طی سال‌های ۱۳۸۴ تا ۱۳۸۸ از بیمارستان ۱۷ شهریور واقع در برازجان مرکز شهرستان دشتستان استان بوشهر، و با توجه به متغیرهای پراهمیت ۱۵متغیر از میان ۶۰ متغیر مرتبط با دیابت افراد دیابتی را در خوشه‌های مختلف قرار داده و براساس تفسیرهای هر خوشه الگوهایی برای افراد دیابتی استخراج کردند [۱۴].

همچنین میرشریف و همکاران، در سال ۱۳۹۶، به دنبال پیش‌بینی ریسک و هشدار به‌موقع در ابتلا به دیابت بارداری به مادر بوده‌اند، تا در اوایل بارداری از ابتلا جلوگیری به عمل آید. این پژوهش که به‌صورت کاربردی‌پیمایشی انجام شد و از ۲ رویکرد شبکه عصبی و درخت تصمیم در داده‌کاوی به‌منظور تجزیه‌وتحلیل آزمایشی داده‌ها و پیش‌بینی استفاده شد [۱۵].

زیبا خوشنامود و همکارانش، پژوهشی با عنوان «تشخیص دیابت به کمک الگوریتم بهینه‌سازی امواج آب و مقایسه آن با الگوریتم‌های یادگیری ماشین» انجام دادند. در این پژوهش یک روش برای تشخیص بیماری دیابت به کمک شبکه عصبی مصنوعی با استفاده از الگوریتم فراابتکاری بهینه‌سازی امواج آب ارائه شد. اطلاعات به‌دست‌آمده از این پژوهش نشان می‌دهد که الگوریتم بهینه‌سازی امواج آب با دقت بیشتری نسبت به ابزارهای داده‌کاوی بیماری دیابت را تشخیص می‌دهد [۱۶].

در تمامی پژوهش‌های انجام‌شده، سیستم‌های تشخیص و پیش‌بینی، قادر به شناسایی بیمار با استفاده از راهکارهای با نظارت هستند. اما در پژوهش پیش‌رو، قصد بر این است که یک رویکرد هوشمند و بدون نظارت، مبتنی بر تکنیک‌های شبکه عصبی‌مصنوعی و الگوریتم ژنتیک جهت تشخیص بیماری دیابت بارداری ارائه شود. سیستم پیشنهادی به دنبال بهبود فرایند تشخیص افراد بیمار از بیمار سالم است. همچنین در سیستم پیشنهادی نیز همانند رویکردهای ارائه‌شده در قبل مسئله دقت تشخیص و پیش‌بینی نیز ملاک ارزیابی است و در اینجا سعی شده است سیستمی ارائه شود که با دقت بالایی بتواند بیماری دیابت بارداری را پیش‌بینی کند.

مواد و روش‌ها

پژوهش حاضر از نوع کاربردی توصیفی است که در سال ۱۳۹۹ انجام شده است. روش پیشنهادی در این پژوهش، یک رویکرد یادگیری بدون نظارت جهت تشخیص دیابت بارداری در افراد مبتلا به این بیماری است. الگوریتم موردنظر، حاصل از آمیختگی الگوریتم ژنتیک و شبکه عصبی‌مصنوعی است. در اینجا ابتدا عملیات خوشه‌بندی بدون نظارت جهت ایجاد ۲ خوشه افراد سالم و بیمار با استفاده از الگوریتم ژنتیک بر روی مجموعه داده ورودی انجام می‌گیرد. درنهایت پس از انجام ۱۰۰ مرتبه اجرا بهترین افراز که حاوی بهینه‌ترین خوشه‌ها به دست می‌آید. بعد از این مرحله، خروجی الگوریتم ژنتیک جهت انجام عملیات آموزش به‌عنوان ورودی شبکه عصبی‌مصنوعی در نظر گرفته می‌شود. شبکه عصبی موردنظر، یک شبکه عصبی پیش‌خور با ۲۰ لایه پنهان است.

پس از طراحی شبکه عصبی پیش‌خور، مجموعه داده ورودی و خوشه‌های به‌دست‌آمده از الگوریتم ژنتیک، که نشان‌دهنده برچسب‌های مجموعه داده است، به‌عنوان ورودی جهت عملیات یادگیری در نظر گرفته می‌شود. در اینجا ۷۵ درصد مجموعه داده به‌عنوان مجموعه داده آموزش و ۲۵ درصد هم به‌عنوان مجموعه داده تست در نظر گرفته شده است. تصویر شماره ۱ شبه‌کد الگوریتم پیشنهادی را نشان می‌دهد.

مجموعه داده مورداستفاده

در این مقاله، از مجموعه داده میرشریف و همکاران استفاده شده است [۱۵]. این مجموعه داده مربوط به بیش از صد پرونده بیماران از سال ۱۳۹۰ تا ۱۳۹۳ در تحقیقی میدانی و نمونه‌گیری هدفمند از کلینیک پزشکی تخصصی زنان در تهران مورد بررسی قرار گرفت که از میان آن‌ها ۱۰۵ مورد دارای اطلاعات کامل از ابتدا تا انتهای بارداری بودند (۸۰ نفر انسان سالم و ۲۵ نفر انسان بیمار).

محیط شبیه‌سازی

در این مقاله، برای بررسی سیستم پیشنهادی و ارائه نتایج آن، پیاده‌سازی رویکرد پیشنهادشده به کمک زبان‌های‌برنامه‌نویسی انجام شده است. به این منظور، الگوریتم پیشنهادشده در محیط متلب نسخه R2016a پیاده‌سازی شده است. زبان برنامه‌نویسی متلب یکی از زبان‌های مفسری است که به‌طور گسترده‌ای در رشته‌های مهندسی استفاده‌می‌شود. این محیط شبیه‌سازی شهرت خود را از قابلیت استفاده آسان خود به دست آورده است. آزمایشات در سیستمی با پردازنده INTEL CELERON D800 1.50 GHz و حافظه‌ی داخلی ۲ گیگابایت مورد تست و ارزیابی قرار گرفته است

```
Input: the set of data items D={x1, x2... xn}, the number of cluster=2.
Generate initial population of size μ, by Run Kmeans and return Centroid.
Evaluate each individual by Run silhousette with;
While (Termination candidate is not satisfied) do
Select parents;
Crossover pairs of parents; /* Crossover rate */
Mutate the resulting offspring; /* Mutation rate */
Evaluate each new offspring by silhousettewth;
/* Now we have λ offsprings */
Select individual for the next generation;
End While.
The Best Solution, a hard partition C= {c1, c2}, U^2_i=c1 Uc2=D and c1=∩ c2=ø
.net=a feed-forward back propagation network;
TRD=create train Data(D);
TSD=create Tes t Data(D);
.net = train(net,TRD,C);
.Output=net(TSD);
Output=the predict of patients.
```

تصویر ۱. شبه‌کد الگوریتم پیشنهادی

جدول ۱. تنظیم مقادیر پارامتر برای الگوریتم پیشنهادی

شرح پارامتر	مقدار
تعداد تکرار	۱۰۰
اندازه جمعیت	۱۰۰
نرخ تقاطع	۰.۷
نرخ جهش	۰.۳
نوع شبکه عصبی	شبکه عصبی پیش‌خور
تعداد لایه‌های پنهان	۲۰
تابع تبدیل	تابع سیگموئیدی
نرخ یادگیری	۷۵ درصد
نرخ تست	۲۵ درصد

پارامترها

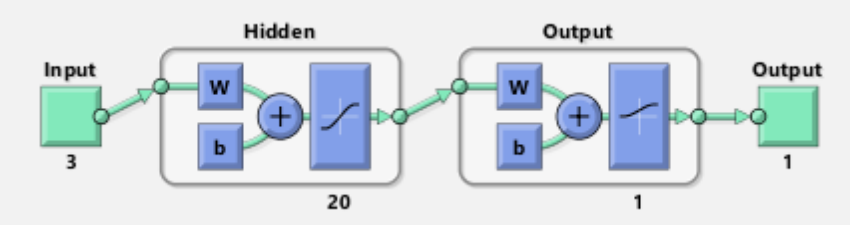
در الگوریتم پیشنهادی مطابق جدول شماره ۱، تعداد تکرار ۱۰۰ مرتبه، تعداد جمعیت اولیه ۱۰۰ عدد، درصد عملیات تقاطع ۷ و درصد عملیات جهش ۳ در نظر گرفته شده است. همچنین این الگوریتم با اجرای یک چرخ رولت عضوهای جمعیت را جهت عملیات تقاطع و جهش انتخاب می‌کند. پس از اجرای عملیات‌های تقاطع و جهش، فرزندهای به‌دست‌آمده از این عملیات‌ها با استفاده از تابع ارزیابی بررسی می‌شوند. سپس این فرزندان با جمعیت قبلی ادغام شده و ۱۰۰ فرد از این جمعیت که دارای مقدار ارزیابی بهتری هستند به‌عنوان جمعیت جدید انتخاب می‌شوند. پس از صد تکرار فردی در جمعیت که دارای بهترین مقدار ارزیابی است به‌عنوان بهترین فرد شناخته شده و پارتشین به‌دست‌آمده از آن به‌عنوان جواب مسئله در نظر گرفته می‌شود که این پارتیشن شامل خوشه‌هایی از مجموعه داده ورودی است. سپس برچسب‌های خوشه‌های به‌دست‌آمده و مجموعه داده به‌عنوان ورودی جهت عملیات یادگیری وارد شبکه عصبی پیش‌خور می‌شود. در شبکه عصبی طراحی‌شده، تعداد لایه‌های پنهان ۲۰، نوع تابع تبدیل تابع سیگموئیدی و نرخ یادگیری و تست نیز به ترتیب ۷۵ و ۲۵ درصد تنظیم شده است.

یافته‌ها

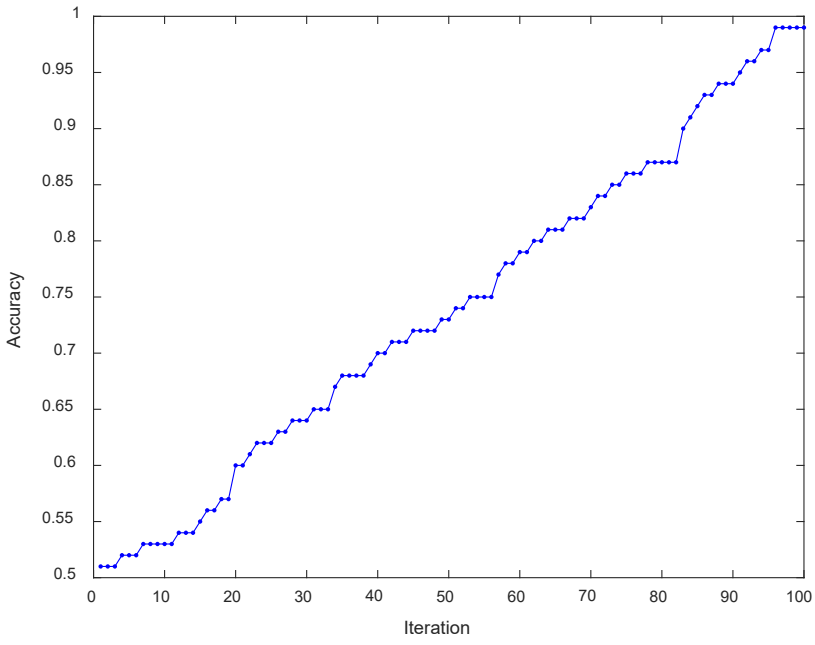
نحوه انجام کار روش پیشنهادی

در این بخش، نحوه انجام کار روش پیشنهادی در محیط متلب توضیح داده می‌شود. در اینجا ابتدا عملیات خوشه‌بندی بدون نظارت جهت ایجاد ۲ خوشه افراد سالم و بیمار با استفاده از الگوریتم ژنتیک بر روی مجموعه داده ورودی انجام می‌گیرد. سپس مجموعه داده ورودی و خروجی الگوریتم ژنتیک جهت انجام عملیات آموزش به‌عنوان ورودی شبکه عصبی مصنوعی طراحی‌شده در نظر گرفته می‌شود. شبکه عصبی‌مصنوعی موردنظر مطابق با تصویر شماره ۲، شبکه عصبی پیش‌خور با ۳ ورودی، ۲۰ لایه پنهان و ۱ خروجی است. همچنین تابع مورد استفاده در آن نیز تابع سیگموئیدی است.

در این تحقیق، به‌طور پیش‌فرض جهت یادگیری سیستم ۷۵ درصد از مجموعه‌ی داده‌ها به‌عنوان نمونه‌های یادگیری مورد استفاده قرار گرفت. همچنین به‌منظور بررسی کردن دقت سیستم از ۲۵ درصد مجموعه داده به‌عنوان تست و اعتبارسنجی استفاده شده است. مجموعه داده موردبررسی، مجموعه داده دیابت استفاده‌شده در مقاله میرشرف و همکاران بود که این مجموعه داده شامل ۱۰۵ عضو بود (۸۰ نمونه افراد سالم و ۲۵ نمونه افراد بیمار) .



تصویر ۲. شبکه عصبی پیش‌خور در طرح پیشنهادی



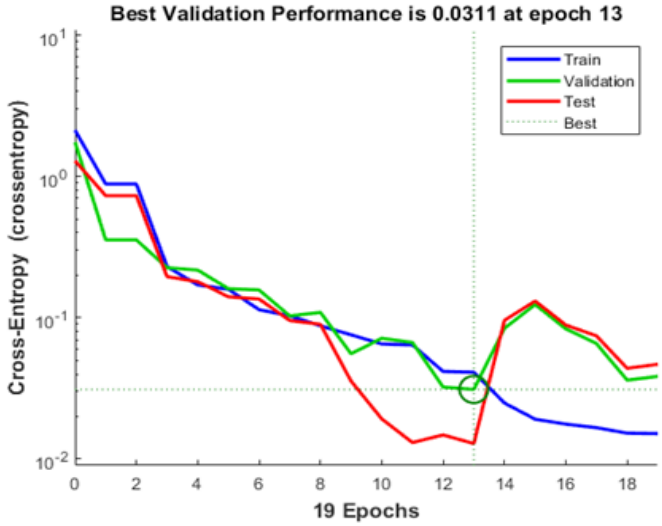
تصویر ۳. نحوه بهبود دقت آموزش سیستم طی ۰۰۱ تکرار

بررسی نحوه بهبود دقت آموزش سیستم

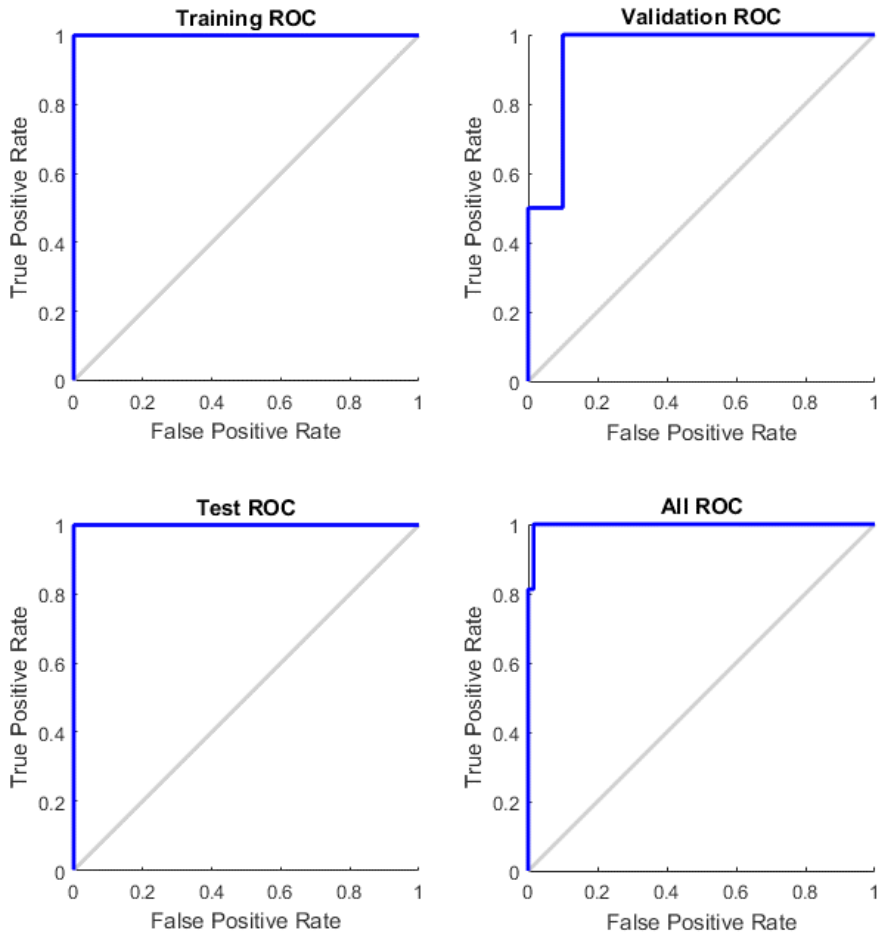
هدف از طراحی این آزمایش بررسی نحوه بهبود دقت سیستم است. برای انجام این کار سیستم با جمعیت ۱۰۰، تکرار ۱۰۰، ضریب تقاطع ۰/۷ و ضریب جهش ۰/۳ اجرا شده است. همچنین در شبکه عصبی طراحی‌شده، تعداد لایه پنهان ۲۰، ورودی ۳ و خروجی ۱ است. در این آزمایش داده‌های تست سیستم (برای ارزیابی سیستم و به دست آوردن دقت) با داده‌های آموزش یکسان است و آزمایش موردنظر ۱۰۰ مرتبه تکرار شده است. در این آزمایش نمودار بهبود در طی تکرارهای متوالی در متلب رسم شده است. همان‌طور که در تصویر ۳ نشان داده شده است. سیستم پیشنهادی توانسته است بعد از گذشت ۱۰۰ تکرار به دقت ۰/۹۸ دست پیداکند. برای به دست آوردن نمودار هموارتر می‌توان سیستم را به تعداد دفعات بیشتر اجرا کرد و از نتایج دفعات مختلف میانگین گرفت و درنهایت روند بهبود را نمایش داد.

بررسی عملکرد سیستم با استفاده از داده‌های آموزش

در این بخش ماتریس درهم‌ریختگی و منحنی راک سیستم بعد از آموزش سیستم بررسی خواهد شد. برای انجام این کار، سیستم پیشنهادی با جمعیت ۱۰۰، تکرار ۱۰۰، ضریب تقاطع



تصویر ۴. میانگین مربعات خطا در الگوریتم پیشنهادی با ۰۲ لایه مخفی



تصویر ۵. منحنی راک مربوط به یادگیری سیستم پیشنهادی

۰/۷، ضریب جهش ۰/۳ و اندازه مجموعه داده ۸۰ نمونه، آموزش دیده است. در این آزمایش از نمونه‌های آموزش، ۶۵ نمونه متعلق به افراد سالم و ۱۵ نمونه متعلق به افراد بیمار بود. تصویر شماره ۴ بهترین کارایی به‌دست‌آمده سیستم پیشنهادی را نشان می‌دهد. مطابق با این تصویر بهترین کارایی ۰/۰۳۱۱ است. هر اندازه این مقدار پایین‌تر باشد نشان از کارایی بالای سیستم است.

در تصویر شماره ۵ منحنی راک[1] نشان داده شده است. راک معیاری است که برای بررسی کیفیت دسته‌بندی‌کننده‌ها مورد استفاده قرار می‌گیرد. برای هر کلاس دسته‌بندی‌کننده، مقادیر آستانه‌ای بین صفر و ۱ برای خروجی‌ها به کار می‌رود. برای هر حد آستانه ۲ مقدار نرخ مثبت واقعی[2] و نرخ مثبت کاذب[3] محاسبه می‌شود. در تصویر شماره ۵ مشاهده می‌شود که همواره نرخ مثبت واقعی بیشتر از نرخ مثبت کاذب است.

در تصویر شماره ۶، ماتریس درهم‌ریختگی مربوط به یادگیری سیستم پیشنهادی نمایش داده شده است. در ماتریس اغتشاش سطرها نشانگر کلاس پیش‌بینی شده (کلاس خروجی) و ستون‌ها نشان‌دهنده کلاس صحیح (کلاس هدف) هستند. سلول‌های مورب نشان می‌دهد چند مقدار (و چه درصدی) از مشاهدات، نمونه شبکه آموزش‌دیده را به‌درستی تخمین زده است.

در تصویر شماره ۶، ۲ سلول مورب اولیه تعداد و درصد دسته‌بندی صحیح با استفاده از شبکه آموزش‌دیده را نشان می‌دهند. در این رقم، ۲ سلول مورب اول، تعداد و درصد طبقه‌بندی‌های صحیح را توسط شبکه آموزش‌دیده نشان می‌دهند. مطابق با نتایج، ۶۴ نفر به‌درستی به‌عنوان افراد سالم طبقه‌بندی شده‌اند. این مطلب مطابق با ۸۰ درصد کل نمونه‌هاست. به‌طور مشابه، ۱۵ مورد به‌درستی به‌عنوان افراد بیمار طبقه‌بندی شده‌اند. این مطلب، مطابق با ۱۸/۸ درصد کل نمونه‌های آزمایشی است. ۱ نمونه از افراد سالم به‌اشتباه بیمار دسته‌بندی شده‌اند و مطابق با ۱/۳ درصد کل ۸۰ نمونه موجود است.

1. Receiver Operating Charactristic
2. True Posetive Rate (TPR)
3. False Positive Rate (FPR)

تصویر ۶. ماتریس درهم‌ریختگی یادگیری سیستم پیشنهادی

سیستم پیشنهادی، مطابق با ماتریس اغتشاش تولیدشده، از ۶۵ پیش‌بینی افراد سالم، ۹۸/۵ درصد به‌درستی و ۱/۵ درصد به‌اشتباه کلاس‌بندی شده‌اند. از ۱۵ پیش‌بینی افراد بیمار، ۱۰۰ درصد به‌درستی کلاس‌بندی شده‌اند. از ۶۴ موارد افراد سالم، ۱۰۰ درصد به‌خوبی به‌عنوان افراد سالم پیش‌بینی شده است. از ۱۶ مورد افراد بیمار، ۹۳/۸ درصد به‌درستی به‌عنوان بیمار و ۶/۳ درصد به‌عنوان افراد سالم طبقه‌بندی شده‌اند. به‌طورکلی، ۹۸/۸ درصد از پیش‌بینی‌ها درست هستند و ۱/۲ درصد طبقه‌بندی‌های اشتباه هستند.

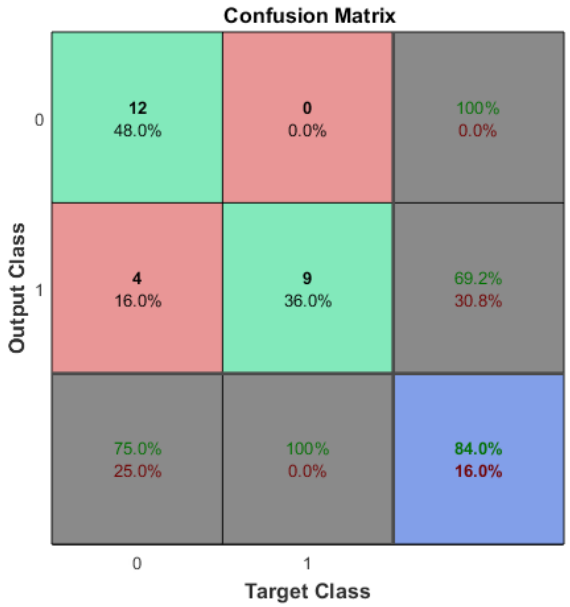
بررسی عملکرد سیستم با استفاده از داده‌های تست

در این آزمایش، شبکه عصبی آموزش‌دیده آزمایش قبل، با داده‌های تست ارزیابی شده است. در این آزمایش داده‌های تست سیستم با داده‌های آموزش یکسان نیست. در این تست از ۲۵ نمونه برای ارزیابی سیستم استفاده شده است که از میان آن‌ها ۱۵ نمونه افراد سالم و ۱۰ نمونه افراد بیمار بودند. در تصویر شماره ۷ ماتریس درهم‌ریختگی به ازای نمونه‌های تست ارائه شده است.

همان‌طور که نمودار ماتریس درهم‌ریختگی نشان می‌دهد، سیستم توانسته به دقت تست ۸۴ درصد دست پیدا کند. همچنین تعداد نمونه‌های افراد سالم نسبت به نمونه‌های افراد بیمار بهتر پیش‌بینی شده‌اند. یکی از دلایل این اتفاق می‌تواند وجود تعداد نمونه‌های مثبت بیشتر در داده‌های آموزش باشد.

مطابق با نتایج ماتریس اغتشاش در فرایند تست، تعداد ۱۲ نمونه به‌درستی به‌عنوان افراد سالم طبقه‌بندی شده است. این مطلب مطابق با ۴۸ درصد کل ۲۵ نمونه است. به‌طور مشابه، ۹ مورد به‌درستی به‌عنوان افراد بیمار طبقه‌بندی شده‌اند. این مطلب، مطابق با ۳۶ درصد کل نمونه‌های آزمایشی است. هیچ‌کدام از افراد سالم به اشتباه در افراد بیمار دسته‌بندی نشده‌اند. اما ۴ مورد از افراد بیمار در کلاس سالم قرار گرفته‌اند که مطابق با ۱۶ درصد کل نمونه‌ها است.

در تست سیستم پیشنهادی، مطابق با ماتریس اغتشاش تولیدشده، از ۱۲ پیش‌بینی افراد سالم، ۱۰۰ درصد به‌درستی کلاس‌بندی شده‌اند. از ۱۳ نمونه پیش‌بینی افراد بیمار، ۶۹.۲ درصد به‌درستی و ۳۰/۸ درصد به‌اشتباه کلاس‌بندی شده‌اند. از ۱۶ نمونه افراد سالم، ۷۵ درصد به‌درستی به‌عنوان افراد سالم و ۲۵ درصد به‌عنوان افراد بیمار پیش‌بینی شده است. از ۹ مورد افراد بیمار، ۱۰۰ درصد به‌درستی به‌عنوان بیمار طبقه‌بندی شده‌اند. به‌طورکلی، ۸۴ درصد از پیش‌بینی‌ها درست هستند و ۱۶ درصد طبقه‌بندی‌های اشتباه هستند.



تصویر ۷. ماتریس درهم‌ریختگی سیستم پیشنهادی با استفاده از داده‌های تست

جدول ۲. مقایسه عملکرد سیستم پیشنهادی و سیستم میرشریف بر روی مجموعه‌داده موردنظر

الگوریتم‌های موردمقایسه	خطای میانگین مربعات	دقت
میرشریف	۰/۰۷۳۸	۰/۹۰۶
سیستم پیشنهادی	۰/۰۰۸۲۳	۰/۹۳۲

مقایسه با تحقیقات انجام‌شده در سال‌های اخیر

در این بخش سیستم پیشنهادی بر روی مجموعه‌داده مشابه، با سیستم پیشنهادشده میرشرف و همکاران جهت تشخیص دیابت بارداری ارزیابی شده است. در این آزمایش سیستم پیشنهادی و سیستم ارائه‌شده توسط میرشریف بر روی مجموعه‌داده مشابه (مجموعه داده مورداستفاده توسط میرشریف) اجراشده و کارایی آن با استفاده از متریک‌های دقت[۴] [۱۵، ۱۷] و میانگین مربعات خطا[۵] [۱۵، ۱۸، ۱۹]، ارزیابی شده است. فرمول شماره ۱ و ۲ معیارهای موردارزیابی را نشان می‌دهد [۱۵].

1. $Accuracy=\frac{TP+TN}{TP+TN+FP+FN}$

2. $MSE=\frac{1}{n}\sum_{i}^{n}=1(Y_i-Y_i')^2$

که در این روابط مقادیر TP، FP و FN به ترتیب برابر با تعداد نمونه‌های مثبت درست پیش‌بینی‌شده، تعداد نمونه‌های مثبت نادرست پیش‌بینی‌شده و تعداد نمونه‌های منفی نادرست پیش‌بینی‌شده هستند. همچنین Y_i مقدار واقعی و Y_i' مقدار پیش‌بینی است. درحقیقت معیار دقت، دقت سیستم موردنظر و معیار خطای میانگین مربعات[۶] میانگین اختلاف مقدار واقعی و مقدار پیش‌بینی‌شده توسط سیستم را به دست می‌آورند. در اینجا، معیار دقت هر اندازه بیشتر باشد سیستم بهینه خواهد بود. اما معیار میانگین مربعات خطا هر اندازه کمتر باشد، سیستم بهینه خواهد بود [۲۰، ۲۱].

در این بخش سیستم پیشنهادی بر روی مجموعه‌داده میرشریف اعمال شده است. مطابق با جدول شماره ۲، کارایی و عملکرد روش پیشنهادی از روش ارائه‌شده توسط میرشریف در هر دو معیار خطای میانگین مربعات و دقت[۷] بهتر است. این نشان می‌دهد که روش پیشنهادی در حالت بدون نظارت کارایی مطلوب را در مواجهه با مجموعه داده‌های مختلف به‌منظور تشخیص دقیق دیابت داراست.

4. Accuracy
5. Mean Square Error
6. Mean Squared Error (MSE)
7. Accuracy (ACC)

بحث

در این پژوهش رویکردی هوشمند و بدون نظارت، جهت تشخیص بیماری دیابت بارداری ارائه شد. روش پیشنهادی به دنبال بهبود فرایند پیش‌بینی افراد داری دیابت بارداری از افراد سالم بود. در همین راستا، زاهدی‌فرد و همکاران، با استفاده از داده‌کاوی، افراد دیابتی را خوشه‌بندی و براساس تفسیرهای هرخوشه الگوهایی برای افراد دیابتی استخراج کردند [۱۴].

همچنین میرشریف و همکاران، سطح هشدار به‌موقع در ابتلا به دیابت بارداری در مادران را بررسی کردند. در این پژوهش با استفاده از شبکه عصبی و درخت تصمیم، روشی جهت پیش‌بینی پیشنهاد شد [۱۵].

زیبا خوشنامود و همکارانش نیز پژوهشی با عنوان «تشخیص دیابت به کمک الگوریتم بهینه‌سازی امواج آب و مقایسه آن با الگوریتم‌های یادگیری ماشین» انجام دادند. در این پژوهش یک روش برای تشخیص بیماری دیابت به کمک شبکه عصبی مصنوعی با استفاده از الگوریتم فرااِبتکاری بهینه‌سازی امواج آب ارائه شد [۱۶].

نتیجه‌گیری

در تمامی این پژوهش‌ها، روش‌های ارائه‌شده همگی قادر به تشخیص و پیش‌بینی، بیماران دیابتی با استفاده از راهکارهای بانظارت هستند. اما در واقعیت، راهکارهای با نظارت نمی‌توانند الگوهای مناسب پیش‌بینی را کشف کنند [۲۲، ۲۳]. به همین منظور، در این پژوهش سعی شده، رویکردی هوشمند و بدون نظارت، مبتنی بر تکنیک‌های شبکه عصبی‌مصنوعی و الگوریتم ژنتیک جهت پیش‌بینی بیماران دیابت بارداری ارائه شود.

شبکه عصبی موجود در روش پیشنهادی الگوهای پنهان در افراد بیمار را نسبت به افراد سالم با بررسی مجموعه داده موردنظر کشف می‌کند. ازآنجاکه نحوه عملکرد شبکه عصبی به چگونگی تعریف ساختار آن بستگی دارد [۲۴، ۲۵]، در این تحقیق در کنار شبکه عصبی پیشنهادشده، از الگوریتم ژنتیک استفاده شده است که وظیفه آن یادگیری بدون نظارت ساختار شبکه عصبی است.

سیستم پیشنهادی در محیط متلب پیاده‌سازی و عملکرد آن ارزیابی شده است. نتایج پیاده‌سازی سیستم پیشنهادی نشان داده است که عملکرد و دقت الگوریتم در مجموعه داده موردنظر و همچنین نسبت به سیستم مشابه قبلی، قابل‌قبول بوده و به‌طور قابل‌توجهی بهبود پیدا کرده است. علاوه بر این نتایج مقایسه سیستم پیشنهادی با روش موجود در حوزه تحقیق نشان داد، سیستم پیشنهادی توانسته است در معیارهای مختلف به بهبود قابل‌ملاحظه‌ای دست یابد.

ازجمله محدودیت‌های موجود در این پژوهش، می‌توان به مدل یادگیر موردنظر اشاره کرد. مدل یادگیر شبکه عصبی به‌شدت وابسته به نوع داده دسته‌بندی‌شده ورودی است. در اینجا، از الگوریتم ژنتیک برای دسته‌بندی داده ورودی استفاده شده است، ازآنجایی‌که یکی از محدودیت‌های الگوریتم ژنتیک، شناسایی دقیق نقطه بهینه در فضای جواب است، پس می‌توان ایجاد دسته‌بندی دقیق مجموعه داده ورودی را به‌عنوان محدودیت روش پیشنهادی دانست. همچنین به‌عنوان پیشنهاد کارهای آتی نیز می‌توان به جایگزین کردن دیگر الگوریتم‌های تکاملی نظیر الگوریتم بهینه‌سازی ازدحام ذرات و الگوریتم کلونی مورچگان به‌جای الگوریتم ژنتیک، در جهت بهبود فرایند دسته‌بندی مجموعه داده ورودی اشاره کرد.

ملاحظات اخلاقی

پیروی از اصول اخلاق پژوهش

این مطالعه دارای تأییدیه اخلاقی به شماره ۱۱۴۴۸۱۹۹۱۶۰۰۴۸۸۱۳۹۷۱۸۷۶۵۹۵۵۶ از دانشگاه آزاد اسلامی واحد استهبان است.

حامی مالی

این مقاله از طرف هیچ نهاد یا مؤسسه‌ای حمایت مالی نشده است و منابع مالی از طرف نویسنده تأمین شده است.

مشارکت نویسندگان

محمدجوادحسین‌پور نویسنده مسئول این مقاله، تمامی مراحل تحقیق و تدوین این مقاله را انجام داده است.

تعارض منافع

بنابر اظهار نویسنده این مقاله تعارض منافع ندارد.

تشکر و قدردانی

نویسنده مقاله بر خود لازم می‌داند از همکاری و مساعدت دانشگاه آزاد اسلامی واحد استهبان سپاس‌گزاری کند.

References

[1] Mierzyński R, Poniedziałek-Czajkowska E, Dłuski D, Patro-Małysza J, Kimber-Trojnar Ż, Majsterek M, et al. Nesfatin-1 and vaspin as potential novel biomarkers for the prediction and early diagnosis of gestational Diabetes Mellitus. Int J Mol Sci. 2019; 20(1):159. [PMID]

[2] Pei T, Wang W, Zhang H, Ma T, Du Y, Zhou C. Density-based clustering for data containing two types of pointsInt J Geogr Inf Sci. 2015; 29(2):175-93. [DOI:10.1080/13658816.2014.955027]

[3] Jayalakshmi T, Santhakumaran A. A novel classification method for diagnosis of Diabetes Mellitus using artificial neural networks. Paper presented at: International Conference on Data Storage and Data Engineering. 9-10 February 2010; Bangalore, India. [DOI:10.1109/DSDE.2010.58]

[4] Kumar D, Palaniswami M, Rajasegarar S, Leckie C, Bezdek JC, Havens TC. clusiVAT: A mixed visual/numerical clustering algorithm for big data. Paper presented at: IEEE International Conference on Big Data. 6-9 October 2013; Silicon Valley, USA. [DOI:10.1109/BigData.2013.6691561]

[5] Zhao L, Ren Y. A scalable genetic algorithm for discovering comprehensible anomaly detection rules using big data in computer cluster. Paper presented at: 3rd International Conference on Systems and Informatics (ICSAI). 19-21 November 2016; Shanghai, China. [DOI:10.1109/ICSAI.2016.7811048]

[6] Johnson T, Kumar Singh S. Enhanced K Strange points clustering algorithm. Paper presented at: International Conference on Emerging Information Technology and Engineering Solutions. 20-21 February 2015; Washington, DC, United States. [DOI:10.1109/EITES.2015.14]

[7] Schneider AK, Leemaqz SY, Dalton J, Verburg PE, Mol BW, Dekker GA, et al. The interaction between metabolic syndrome and physical activity, and risk for gestational Diabetes Mellitus. Acta Diabetol. 2021; 58(7):939-47. [PMID]

[8] Berikov V. Weighted ensemble of algorithms for complex data clustering. Pattern Recognit Lett. 2014; 38:99-106. [DOI:10.1016/j.patrec.2013.11.012]

[9] Khan SR, Mohan H, Liu Y, Batchuluun B, Gohil H, Al Rijjal D, et al. Correction to: The discovery of novel predictive biomarkers and early-stage pathophysiology for the transition from gestational diabetes to type 2 diabetes. Diabetologia. 2019; 62(4):730-1. [PMID]

[10] Breuing J, Pieper D, Neuhaus AL, Heß S, Lütkemeier L, Haas F, et al. Barriers and facilitating factors in the prevention of diabetes type II and gestational diabetes in vulnerable groups: protocol for a scoping review. Syst Rev. 2018; 7(1):245. [PMID]

[11] Parsons J, Sparrow K, Ismail K, Hunt K, Rogers H, Forbes A. Experiences of gestational diabetes and gestational diabetes care: A focus group and interview study. BMC Pregnancy Childbirth. 2018; 18(1):25. [PMID]

[12] Hosseinpoor M, Parvin H, Nejatian S, Rezaei V. [Detection and extraction of potential promoter/enhancer interactions in genome of cancer patients using an evolutionary multi-objective algorithm (Persian)]. J Health Biomed Inform. 2018; 5(2):304-13. [Link]

[13] Mohammadi F, Nazari S. [Exodite segmentation in diabetic patients in retinal images using fuzzy clustering (Persian)]. Papaer presented at: 3rd International Congress on Computer, Electrical and Telecommunication. 22 September 2016; Torbat Heydariyeh, Iran. [Link]

[14] Zahedi Fard MR, Malekzadeh J, Habibi S. [Medical data mining: Pattern discovery for diabetics using significant variables in diabetes (Persian)]. Paper presented at: 12th National Conference on Intelligent Systems, Bam, Iranian Intelligent Systems Association. 3-5 February 2014; Bam, Iran. [Link]

[15] Mirsharif M, Rouhani S. [Data mining approach based on neural network and decision tree methods for the early diagnosis of risk of gestational Diabetes Mellitus (Persian)]. J Health Biomed Inform. 2017; 4(1):59-68. [Link]

[16] Khoshnamvand Z, Asadi F, Khoshnamvand S, Khoshnamvand M. Diagnosis of diabetes using water wave optimization algorithm and comparison with machine learning algorithms. Paper presented at: 5th International Conference on Knowledge Based Research in Computer Engineering and Information Technology. 21 July 2017; Tehran, Iran. [Link]

[17] Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, et al. IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract. 2022; 183:109119. [PMID]

[18] Pouya S, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9(th) edition. Diabetes Res Clin Pract. 2019; 157:107843. [DOI:10.1016/j.diabres.2019.107843]

[19] Stotz SA, McNealy K, Begay RL, DeSanto K, Manson SM, Moore KR. Multi-level diabetes prevention and treatment interventions for native people in the USA and Canada: A scoping review. Curr Diab Rep. 2021; 21(11):46. [PMID]

[20] Teufel NI, Ritenbaugh CK. Development of a primary prevention program: insight gained in the Zuni Diabetes Prevention Program. Clin Pediatr (Phila). 1998; 37(2):131-41.[PMID]

[21] Chambers RA, Rosenstock S, Neault N, Kenney A, Richards J, Begay K, et al. A home-visiting diabetes prevention and management program for American Indian Youth: The together on Diabetes trial. Diabetes Educ. 2015; 41(6):729-47. [PMID]

[22] Hosseinpoor MJ, Parvin H, Nejatian S, Rezaie V. Gene regulatory elements extraction in breast cancer by Hi-C data using a meta-heuristic method. Russ J Genet. 2019; 55(9):1152-64. [DOI:10.1134 / S1022795419090072]

[23] Kulhawy-Wibe S, King-Shier KM, Barnabe C, Manns BJ, Hemmelgarn BR, Campbell DJT. Exploring structural barriers to diabetes self-management in Alberta First Nations communities. Diabetol Metab Syndr. 2018; 10:87. [PMID]

[24] HAPO Study Cooperative Research Group. Hyperglycemia and Adverse Pregnancy Outcome (HAPO) Study: Associations with neonatal anthropometrics. Diabetes. 2009; 58(2):453-9. [PMID]

[25] Marais C, Hall DR, van Wyk L, Conradie M. Randomized cross-over trial comparing the diagnosis of gestational diabetes by oral glucose tolerance test and a designed breakfast glucose profile. Int J Gynaecol Obstet. 2018; 141(1):85-90.[PMID]

This Page Intentionally Left Blank